EXHIBIT A

平成元年5月24日第三種郵便物認可　平成11年12月20日印刷　平成11年12月25日発行　第132号　(毎月1回25日発行)

THE TRANSACTIONS OF THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS D-II

電子情報通信学会 論文誌 D-II

VOL.J82-D-II NO.12 DECEMBER 1999

情報・システムII—パターン処理

論　文

〔パターン認識〕

遺伝的アルゴリズムを用いた対象のユークリッド空間への配置法
……大町真一郎　横山弘子　阿曽弘具　2195

〔音声，聴覚〕

対数スペクトルの任意基底関数による展開に基づく音声のスペクトル推定
……若子武士　徳田恵一　益子貴史　小林隆夫　北村　正　2203

事前に他人受理誤り率を指定する話者照合方式……早川昭二　武田一哉　板倉文忠　2212

〔画像処理，画像パターン認識〕

ハフ変換を用いた接線情報の抽出と欠損楕円の検出……渡辺孝志　畠山雅充　木村彰男　2221

2次元等価近似法による各種変調画像の評価法……川崎順治　飯島泰蔵　2230

ECM法を用いた確率分布の位置，尺度，回転パラメータの推定法……赤穂昭太郎　2240

空間分割モデルを用いた3次元モフォロジーフィルタ
……西尾孝治　小堀研一　久津輪敏郎　西川禕一　2251

自然なテクスチャの特徴抽出用「形状通過型」非線形フィルタバンク
……田村　仁　阿刀田央一　2260

拡張外郭方向寄与度特徴と輪郭特徴とを用いた手書き漢字/非漢字のハイブリッド認識
……木村義政　秋山照雄　森　稔　宮本信夫　若原　徹　小倉健司　2271

高速ビジョンのためのSelf Windowing……石井　抱　石川正俊　2280

シーン仮説と入力画像との大局的画像整合性評価に基づく複数物体の認識
……橋本　学　黒田伸一　鷲見和彦　宇佐美照夫　仲田周次　2288

計測誤差を考慮した距離画像からの精密な姿勢推定……清水郁子　出口光一郎　2298

ネットワーク型並列計算環境における物体認識
……大上靖弘　杉本和英　北村　徹　角　保志　富田文明　2307

領域分割を用いた画像による駐車車両検出法……山田啓一　水野守倫　2316

メタボールを用いた2次元画像符号化方式……増倉孝一　熊澤逸夫　2325

視覚センサと距離センサによる物体認識のための能動的視点制御……守田　了　2335

動的輪郭モデルによるMR画像における左室内腔自動輪郭抽出法の開発とその評価
——輪郭形状の主成分分析を利用した初期値設定——
……湯浅真由美　渡邊　睦　西浦正英　山口弘次郎　近藤　武　安野泰史　武藤晃一　2345

〈裏面へつづく〉

情報・システムソサイエティ
社団法人 電子情報通信学会
THE INFORMATION AND SYSTEMS SOCIETY
THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

信学論(D-II)　Vol. J82-D-II　No. 12　pp. 2195-2454　1999/12/25　東京

論　文

事前に他人受理誤り率を指定する話者照合方式

早川　昭二[†*]　　武田　一哉[†]　　板倉　文忠[††]

A Speaker Verification Method Which can Control False Acceptance Rate

Shoji HAYAKAWA[†*], Kazuya TAKEDA[†], and Fumitada ITAKURA[††]

あらまし　本論文では，事前に他人受理誤り率を指定する照合方式を提案する．本方式では，あらかじめ保持する他人話者群と入力音声との距離から話者間距離の確率分布を照合ごとに推定し，入力音声と自分であると名乗った話者との距離がこの確率分布に含まれる確率を計算して，あらかじめ指定した他人受理誤り率の確率を超えているか否かで話者を照合する．話者間距離の確率分布を正規分布と仮定して照合実験を行った結果，男性15名が最大1年の時期差をおいて発声した単語データベース及び電話回線上で収録した男性70名女性30名の数字音声データベースの両方において，他人受理誤り率は，正規分布と仮定した場合の理論値とほぼ一致し，事前に他人受理誤り率を指定できることを示す．また，提案する照合方式における他人受理誤り率は，発声時期差や雑音による変動の影響をほとんど受けないことを示す．

キーワード　話者照合，他人受理誤り率，話者間距離分布，電話回線，しきい値

1. まえがき

話者認識技術の高度化のため，今後解決しなければならない研究課題として，事前に適切な判定のしきい値を設定する話者照合方法の確立が挙げられている[1]．現在行われている話者照合の研究では，2種類の誤り率：

（1）詐称者の音声を誤って本人の音声として受理してしまう率（False Acceptance rate [FA率]）

（2）本人の音声を誤って詐称者の音声として棄却してしまう率（False Rejection rate [FR率]）

が等しくなるように判定のしきい値を事後的に決めて，そのときの等誤り率（Equal Error Rate）で照合性能を評価することが多い．しかし，実際のフィールドにおいて事前に等誤り率となるしきい値を設定することは非常に難しい．

話者認識における重要な問題として発声時期差による特徴パラメータの変動がある．これは同じ人が同じ内容を発声しても，ある程度長い期間をおくと特徴パラメータが変化してしまい，話者認識性能に影響を与えるというものである[2]．このため話者照合の場合には，登録時とシステム利用時との時期差が大きくなると登録音声との距離が変化してしまう．しかもどのように変化するかは個人によって異なり，事前に予測することは困難である[3]．このことが，等誤り率となるしきい値を事前に設定することを困難にしている．

この問題に対して，古井は学習データにおける本人のテンプレートと本人以外のデータとの類似度の平均と標準偏差を用いてしきい値を決定する方法，つまり比較的推定精度のよいFA率だけを考慮する方法を提案している[4]．松井は話者モデルの更新ごとに等誤り率となるしきい値に次第に近づけていく方法を提案している[5]．しかしいずれの方法においても，照合方式の中に予備実験により決定するパラメータが含まれており，何らかの事後的な情報が利用されている．更に，誤り率を事前に指定してしきい値を決める方法はまだ確立されていない．

本論文では，話者間距離の分布が話者内距離の分布よりも比較的安定に推定できることに注目し，claim話者との距離が話者間距離の分布に含まれる確率を用いて判定することにより，事前に他人受理誤り率を指定して話者を照合する方式を提案する．

† 名古屋大学大学院工学研究科，名古屋市
Graduate School of Engineering, Nagoya University, Nagoya-shi, 464–8603 Japan
†† 名古屋大学情報メディア教育センター統合音響情報研究拠点，名古屋市
Center for Information Media Studies, Center for Integrated Acoustic Information Research, Nagoya University, Nagoya-shi, 464–8603 Japan
* 現在，(株) 富士通研究所パーソナル&サービス研究所

なお本論文では以下，話者照合時に自分であると名乗った話者のことを「claim 話者」(発声者が claim された話者本人とは限らない)，claim された話者以外の登録話者を「他人話者群」と呼ぶことにする．

2. 話者間距離の分布による話者照合方式

本人の声は発声ごとに変化するが，本人と他の話者との相対的な関係は話者内での変動と比して安定であると思われる．松井らは実際に発声時期差のある音声データを用いて，テキスト独立型話者照合実験を行い，FR 率の方が FA 率よりも時期差による発声変動の影響を大きく受けることを示した [5]．この結果は，話者照合において「本人である可能性」を判定するよりも「他人でない可能性」をある危険率で判定する方が発声ごとの変動に対して安定であることを示唆している．

そこで，話者照合時に claim 話者に対する距離だけでなく，他人話者に対する距離を求め，その確率分布を推定することにより，claim 話者が他人話者群に含まれる確率を求め，話者を照合する方式を提案する．

提案する照合方式のブロック図を図 1 に示す．「話者 k である」という claim とともに音声データが入力されると，登録されている N 人すべての話者の登録音声との間で距離計算を行い，各々の話者との距離 $d_i(i = 1, 2, \cdots, N)$ を求める．次に claim された話者との距離 d_k を除いた $N-1$ 個の距離の集合 $\{d_i | i \neq k\}$ から，他人話者群と入力音声との距離の分布 $F(d;\theta)$ を推定する（以降では $F(d;\theta)$ を話者間距離の分布と呼ぶ）．ここで θ は分布のパラメータである．最終的な受理・棄却の判定は，claim された話者との距離 d_k が分布 $F(\cdot;\theta)$ より出力される確率値 $F(d_k;\theta)$ と，あらかじめ設定してある FA 率と比較することで行う．

話者間距離の分布には，混合ガウス分布を用いれば複雑な分布も近似できるが，本論文では取扱いが容易な単一の正規分布で話者間距離の分布を近似する．図 2 に 3. の実験で用いる男性話者の発声した単語データにおける話者間距離の分布を示す．距離は詐称者の入力ごとに，全登録話者に対する距離の平均値と標準偏差で正規化されている(注1)．実線は正規分布の確率密度関数である．この図から話者間距離の分布の概形は，ほぼ正規分布に従っていることがわかる．

今，話者間距離の分布を正規分布で近似した場合，確率分布パラメータである平均値 μ と標準偏差 σ は次式により求められる．

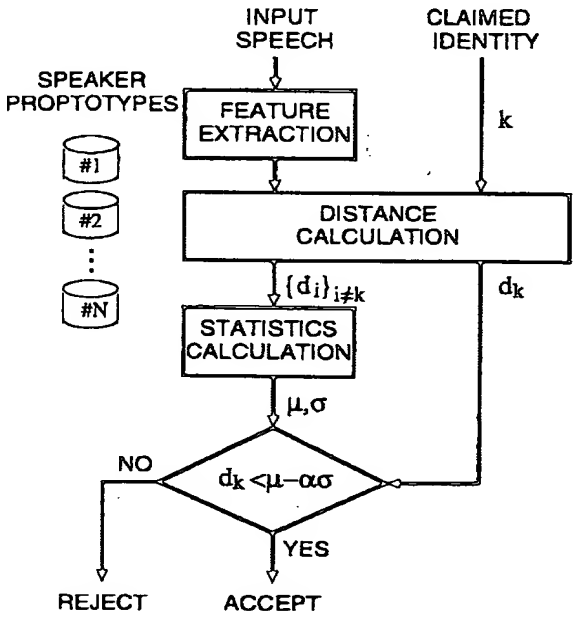

図 1 話者間距離の分布による話者照合方式のブロック図
Fig. 1 Block diagram for proposed speaker verification method.

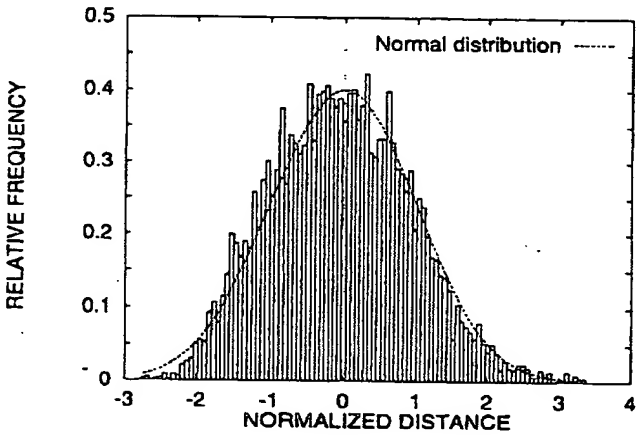

図 2 話者間距離の分布
Fig. 2 Distribution of the interspeaker distances.

$$\mu = \frac{1}{N-1}\sum_{\substack{n=1 \\ n\neq k}}^{N} d_n, \tag{1}$$

$$\sigma^2 = \frac{1}{N-2}\sum_{\substack{n=1 \\ n\neq k}}^{N} (d_n - \mu)^2. \tag{2}$$

(注1)：本論文で提案する話者照合方式に用いる照合距離に相当している．

ただし，claim 話者を k としている．話者間距離の平均値 μ と標準偏差 σ から次の照合式を構成する，

$$\begin{array}{lll} \text{if } d_k & < & \mu-\alpha\cdot\sigma \text{ then accept,} \\ \text{if } d_k & \geqq & \mu-\alpha\cdot\sigma \text{ then reject.} \end{array} \quad (3)$$

α はあらかじめ指定した FA 率に対応して正規確率分布関数

$$\Phi(\alpha)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\alpha}e^{\frac{-t^2}{2}}dt, \quad (4)$$

より求められる正規化距離である．入力された音声が claim 話者本人の音声ならば，その（話者内）距離は話者間距離分布の平均値より必ず小さい値になると考えられるので，正規分布関数の片側だけに含まれる確率で判定する．例えば FA 率を 5%に設定する場合，$1-\Phi(\alpha)=0.05$ すなわち $\alpha=1.65$ を用いる．

古井は話者間距離の分布が話者内距離の分布に比べて比較的安定していることを利用して照合のしきい値を決定する方法を提案している [4]．話者 k のテンプレートが更新されたとき，k を除く $N-1$ 人の話者の発声データ（これはシステムが保持している）と更新された話者 k のテンプレートとの距離の平均値と分散，

$$\mu_k=\frac{1}{N-1}\sum_{\substack{n=1\\ n\neq k}}^{N}d(k,n), \quad (5)$$

$$\sigma_k^2=\frac{1}{N-2}\sum_{\substack{n=1\\ n\neq k}}^{N}(d(k,n)-\mu_k)^2, \quad (6)$$

に基づいて話者 k のしきい値 θ_k を，

$$\theta_k=a(\mu_k-\sigma_k)+b \quad (7)$$

により更新する方法である．ただし $d(k,n)$ は話者 k のテンプレートと話者 n の発声データの間の距離を表す．a と b は定数パラメータですべての話者に対して共通な値を予備実験により設定している．

本論文で提案する手法は話者間の距離の分布を判定に用いる点で古井の提案した方法と類似する一方，以下の二つの特長をもっている．

（1） 本手法は照合すべき claim 話者の入力音声を用いて話者間距離を計算し，その相対的な値により照合を行うことで，cohort モデル [6] を用いる場合と同様に，時期差や発声環境による変動に対して頑健な判定が期待できること，

（2） 古井の提案する手法のように a,b といったパラメータを求めることなく，他人受理率（FA 率）を設定した判定が可能であること．

3. 話者照合実験

提案する照合方式の FA 率と，話者間距離の分布が正規分布に従うと仮定した場合の理論値とを比較する．実験の結果より得られた FA 率と理論値が一致すれば，あらかじめ定めた FA 率をもつ話者照合システムが構築されることになる．また，特徴パラメータが異なる場合や発声時期差及び発声環境の変動に対する影響についても調べる．

3.1 実験条件

データベース及び分析条件を表 1 に示す．各登録話者 1 単語につき 5 個の標準パターンを参照するマルチテンプレート法による発声内容依存型の話者照合実験を行う．ケプストラムと Δ ケプストラムをそれぞれ話者内標準偏差で正規化した後，連結し 28 次元のベクトルとして扱う [7]．距離の計算は対称型擬似端点フリー DTW で行い，最初の収録時の音声を標準パターンとし，残りの 4 時期をテストデータとした [8]．詐称者音声データとしては，登録話者と同じ収録系を用いて，5 単語を 5 回繰り返して発声した音声を 1 回収録したものを用いた．よって本人の照合回数が 1500 回（15 名× 5 単語× 5 回× 4 時期），詐称者の照合回数が 450 回（18 名× 5 単語× 5 回）である．

3.2 15 名の男性話者による話者照合実験の結果

実験結果を図 3 に示す．縦軸は照合誤り率を示し，横軸は式 (3) における α である．$\alpha=0$ の場合は，しきい値を話者間距離の平均値に設定した場合に相当する．$1-\Phi(\alpha)$ は，正規分布関数を 1 から引いた値を示し，話者間距離の分布が正規分布に従っていると仮

表 1 データベースと分析条件
Table 1 Database and analysis conditions.

登録話者数	成人男性話者 15 人
詐称話者数	成人男性話者 18 人
登録話者の収録時期	3 カ月毎に 1 年間
発声単語	爆音，名前，海，高原，番号
単語長	約 0.3〜1 秒
各時期の発声回数	各話者各単語 5 回
標本化周波数 f_s	8 kHz
遮断周波数 f_c	3.4 kHz
フレーム長	32 ms
フレーム周期	16 ms
高域強調	なし
LPC 分析次数	12 次
ケプストラム次数	14 次
Δ ケプストラム窓長	±3 フレーム（112 ms）

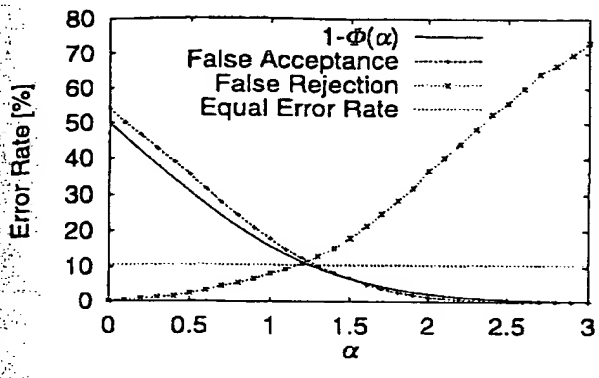

図3 本手法によってしきい値を決定した場合のFA率，FR率及び等誤り率

Fig. 3 False acceptance rate, false rejection rate and equal error rate for the proposed speaker verification method.

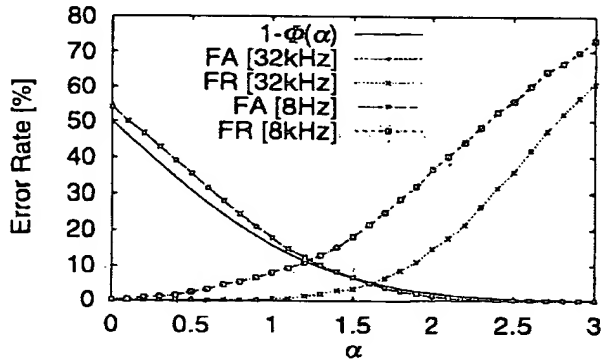

図5 照合性能差がある2種類の特徴パラメータに対するFA率とFR率

Fig. 5 False acceptance rate and false rejection rate for two types of feature parameters which have different speaker verification performance.

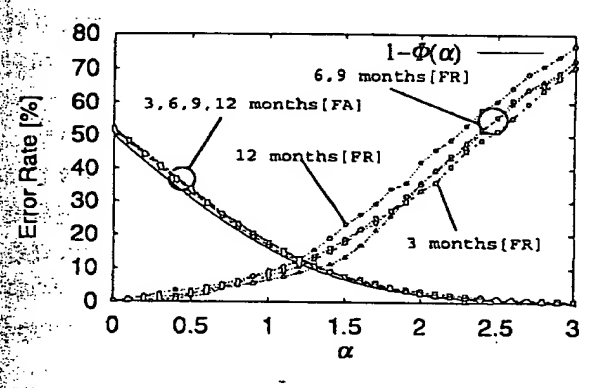

図4 本手法における発声時期ごとのFA率とFR率

Fig. 4 False acceptance rate and false rejection rate for each utterance period.

定した場合のFA率の理論値である．

実験で得られたFA率と理論値がほぼ一致しており，話者間距離分布の正規分布での近似の妥当性が確認された．

3.3 発声時期差による変動の影響

本照合方式の発声時期差に対する耐性を調べる．詐称者データは1時期のみの収録音声のため，本実験では15名の登録話者のうち，1名をclaim話者とし，残りの14名を詐称者とした話者照合を行う．すなわち本実験では，詐称者は他人話者群に含まれる話者である．

図4に時期ごとの結果を示す．FA率は時期差による変動の影響をほとんど受けず，時期が経っても正規分布に従う傾向がある．一方，FR率は発声時期差によって誤り率の傾向が大きく変動する．これらの結果は松井らによる実験結果と一致する[5]．また提案する照合方式は，発声時期差が大きくなっても常に事前に指定したFA率と一致することがわかる．なお，他人話者群に含まれない話者を詐称者として用いた実験（3.2，3.4，3.5）と本節の実験との間で，FA率とαとの関係に顕著な差は見られなかった．

3.4 特徴パラメータの違いによる影響

次に，照合性能が異なる特徴パラメータを用いた場合の，誤り率への影響について調べる．具体的には広帯域の音声データを用い[8]，ダウンサンプリング前の32 kHzサンプリングのデータと前節で用いた8 kHzサンプリングのデータとを比較する．32 kHzサンプリングの場合のLPC分析次数は36次，ケプストラム次数，Δケプストラム次数はそれぞれ14次とし，前節の実験と同様に話者内標準偏差で正規化した後，一つのベクトルとして扱った．

照合実験の結果を図5に示す．照合性能の差は主にFR率に現れる．広帯域の音声を用いた照合性能の高い特徴パラメータでは低いFR率が得られる．一方，FA率の曲線は照合性能にかかわらず理論値に従っている．この結果から，提案する手法では特徴量の照合性能にかかわらず，事前に指定したFA率が得られることが確認された．

3.5 環境の違いによる傾向

本論文で提案する照合方式は発声ごとにclaim話者との距離と他人話者群との距離の相対的な位置関係を用いて照合する．したがってcohortモデルを用いた場合と同様の効果が得られ，発声環境の変動に対して頑健であると考えられる．そこで，登録時と照合時に発声環境が異なる場合の照合性能について検討する．登録音声はクリーンな環境下で発声した音声を用い，照

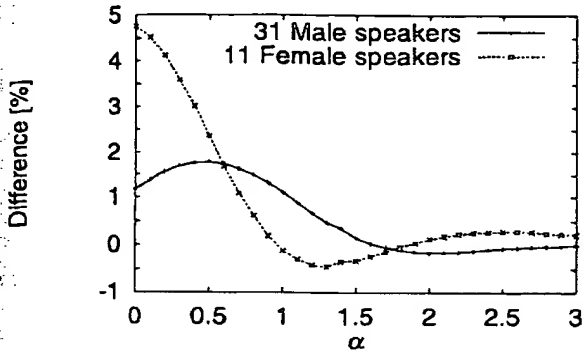

図 12　最も偏差の平均値が小さい場合の各 α に対する理論値と実験値との差
Fig. 12　Difference between theoretical and experimental error rate when using optimal number of speakers for estimating distribution of interspeaker distance.

推定する話者数を増やしても 0 に近づかないことから，話者間距離は完全には正規分布に従っていない．しかし，実際のフィールドで問題となるのは比較的低い FA 率における精度（正規分布の裾）なので，これまでの結果から正規分布として近似しても有効性は損なわれないと考えられる．

6. むすび

本論文では，発声ごとに話者間距離の確率分布を推定し，その分布に含まれる確率をあらかじめ指定した FA 率で評価することによって，利用者本人か否かを照合する方式を提案した．そして，話者間距離の分布を取扱いが容易な正規分布で近似した場合の FA 率の理論値と比較した．

実験の結果，ほぼ理論値に従った FA 率が得られ，本手法によって事前に FA 率を指定できることを示した．また，FA 率の傾向は，発声時期差や認識性能，発声環境の変動の影響を受けにくいことを示した．

また，電話回線上で収録した大量の音声データにおいても同様の結果が得られ，本手法の一般性が示された．また，話者間距離の分布を男女混合で推定する問題点を示し，男女別に推定することで解決できることを示した．またこの方式は，cohort モデルによる正規化と同じように等誤り率を減らす効果があることを示した．

最後に話者間距離の分布の推定に用いる話者の数と FA 率の理論値との誤差ついて調べた結果，10～20 名の他人話者を推定に用いることで理論値との誤差が小さくなることがわかった．

本照合方式は音声による個人認証に限られた方式ではなく，様々な個人性情報を用いた個人認証への応用が考えられる．

謝辞　本研究は文部省科学研究費補助金による研究成果である．電話音声データベースを提供して頂いた国際電信電話株式会社に感謝します．

文　献

[1] 古井貞熙，“話者認識技術の高度化を目指して，” 音学講論，1-7-10, pp.611–614, Oct. 1993.
[2] 古井貞熙，“音声の個人性パラメータの時期的変動と話者認識，” 信学論（A），vol.J57-A, no.12, pp.880–887, Dec. 1974.
[3] 古井貞熙，“音声波に含まれる個人性情報，” 東京大学学位論文，1978.
[4] S. Furui, “Cepstral analysis technique for automatic speaker verification,” IEEE Trans. Acoust., Speech, & Signal Process., vol.ASSP-29, no.2, pp.254–272, April 1981.
[5] 松井知子，西谷　隆，古井貞熙，“話者照合におけるモデルとしきい値の更新法，” 信学論（D-II），vol.J81-D-II, no.2, pp.268–276, Feb. 1998.
[6] A.E. Rosenberg, J. Delong, C.H. Lee, B.H. Juang, and F.K. Soong, “The use of cohort normalized scores for speaker recognition,” Proc. ICSLP, vol.1, pp.599–602, Oct. 1992.
[7] 松井知子，古井貞熙，“音源・声道特徴を用いたテキスト独立型話者認識，” 信学論（A），vol.J75-A, no.4, pp.703–709, April 1992.
[8] 早川昭二，板倉文忠，“音声の高域に含まれる個人性情報を用いた話者認識，” 音響誌，vol.51, no.11, pp.861–868, Nov. 1995.
[9] 梅崎太造，板倉文忠，“平滑化群遅延スペクトル距離尺度の特定話者音声認識における評価，” 信学論（A），vol.J73-A, no.4, pp.734–740, April 1990.
[10] 内部利明，黒岩眞吾，樋口宜男，“数字を用いた話者照合方式の検討，” 信学技報，SP98-68, Oct. 1998.
[11] J. Makhoul, “Spectral linear prediction: Properties and applications,” IEEE Trans. Acoust., Speech and Signal Process., ASSP-23, no.3, pp.283–296, June 1975.
[12] 内部利明，黒岩眞吾，八戸文廣，“話者照合における VQ 歪みを用いた DP 距離正規化法，” 音学講論，1-1-18, pp.35–36, Sept. 1997.

（平成 11 年 2 月 22 日受付，7 月 14 日再受付）